Case Studies 2022L

Partial Dependence Profiles

Apr 21, 2022

Partial dependence profiles

- Friedman (2000) introduced the PD profiles in the context of gradient boosting machines (GBM).
- For many years, PD profiles went unnoticed in the shadow of GBM.
- In recent years, they have become very popular and are available in many data-science-oriented packages like DALEX, iml, pdp, and PDPbox.
- The underlying idea behind the PD profiles is to show how does the expected value of model prediction behave as a function of a selected predictor.
- For a single model, one can construct an overall PD profile by using all observations from a dataset, or several profiles for sub-groups of the observations. Comparison of sub-group-specific profiles may provide important insight into, for instance, the stability of the model's predictions.

Partial dependence profiles

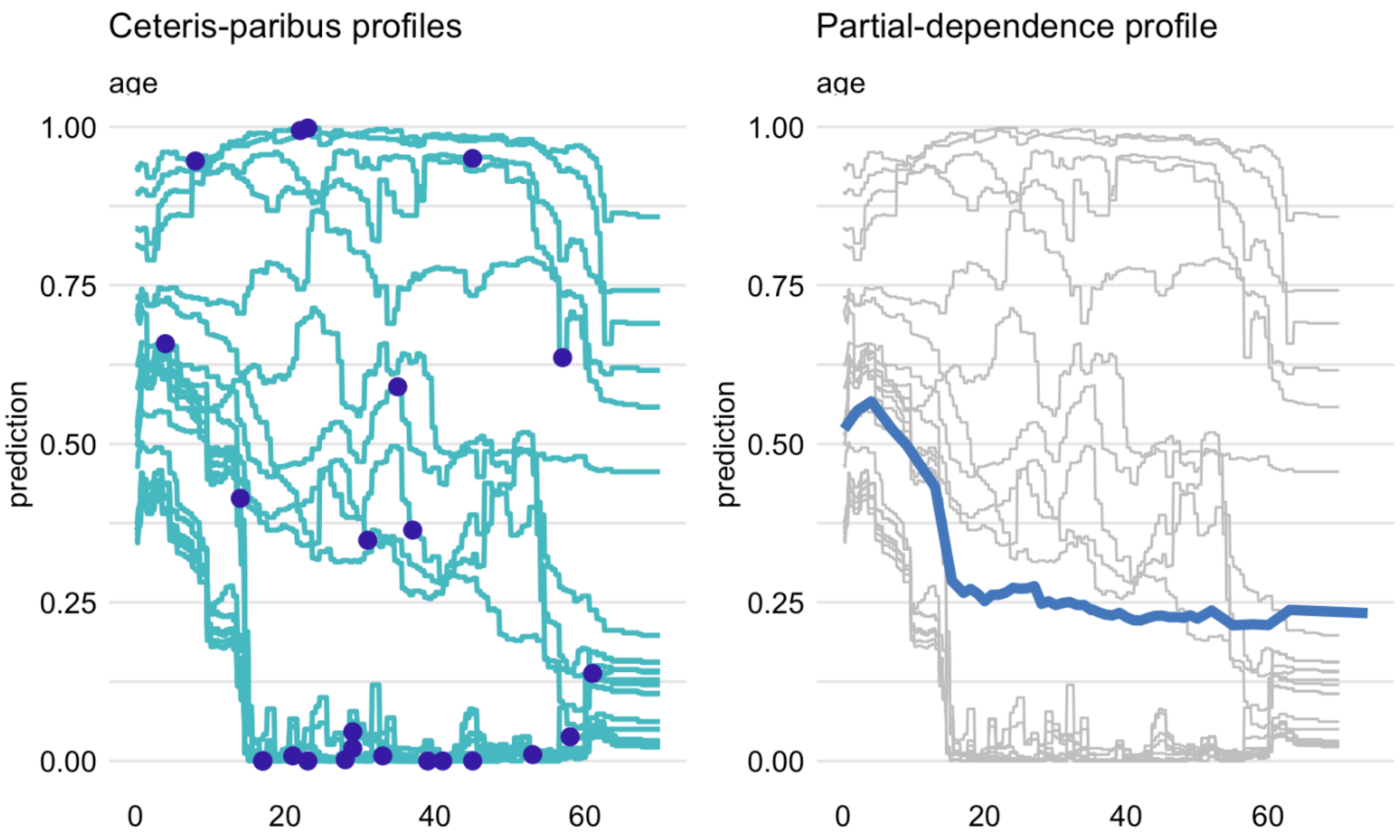
PD profiles are also useful for comparisons of different models:

- Agreement between profiles for different models is reassuring
- Disagreement between profiles may suggest a way to improve a model
- Evaluation of model performance at boundaries

Method

- Recall that a CP profile shows the dependence of an instance-level prediction on an explanatory variable.

- A PD profile is estimated by the mean of the CP profiles for all observations from a dataset.



Ceteris-paribus (CP) and partial-dependence (PD) profiles for the random forest model for 25 randomly selected observations from the Titanic dataset. Left-hand-side plot: CP profiles for *age*; blue dots indicate the age and the corresponding prediction for the selected observations. Right-hand-side plot: CP profiles (grey lines) and the corresponding PD profile (blue line).

Method

The value of a PD profile for model f() and explanatory variable X^j at z is defined as follows:

$$g_{PD}^{j}(z) = E_{\underline{X}^{-j}}[f(X^{j|=z})] = \frac{1}{n} \sum f(\underline{x}_i^{j|=z})$$

Thus, it is the expected value of the model predictions when X^j is fixed at z over the distribution of $\underline{X}^{-j}$ over the joint distribution of all explanatory variables other than X^j.

Types of partial dependence profiles

- Clustered
- Grouped
- Contrastive

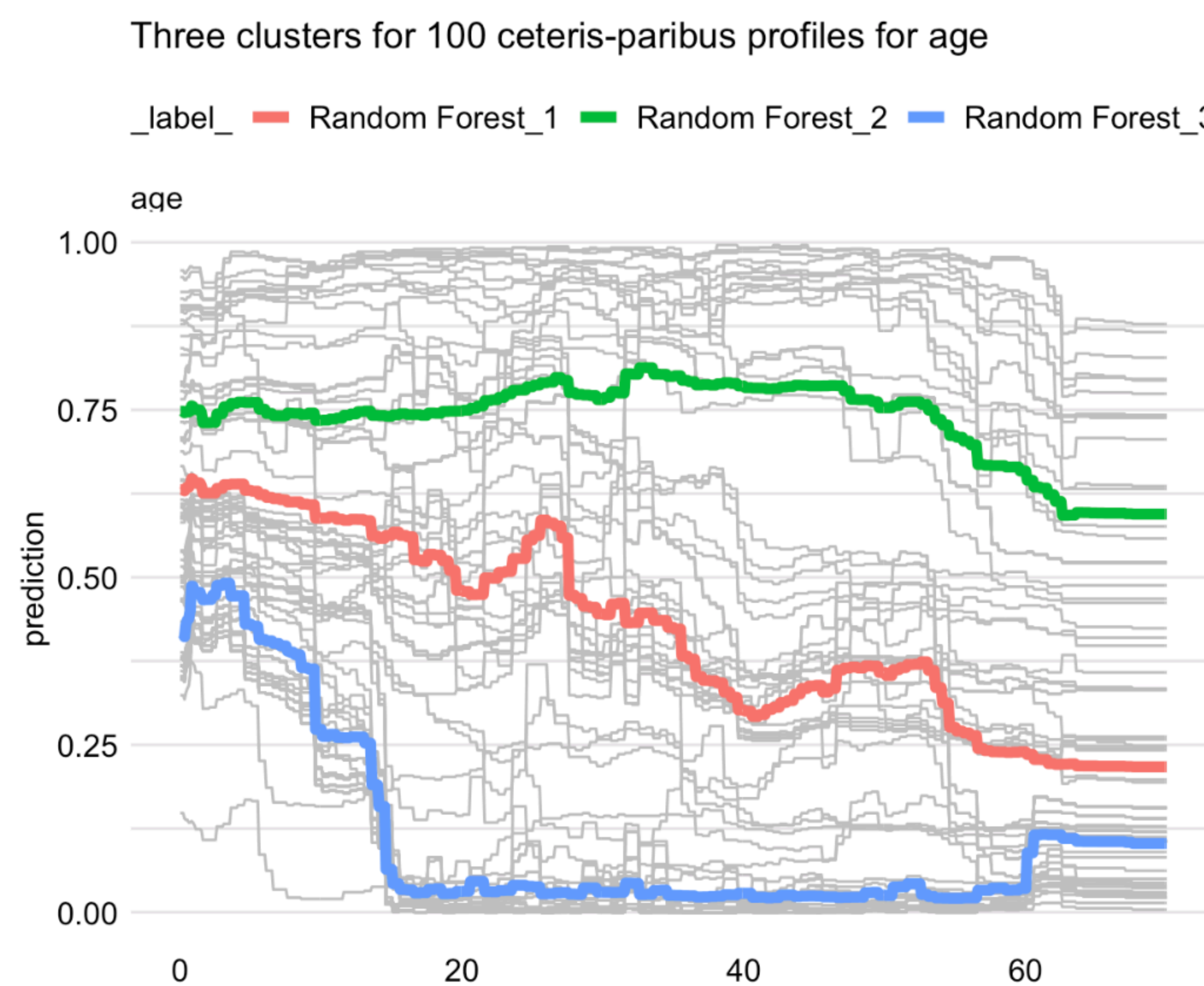
Clustered partial-dependence profiles

- The mean of CP profiles is a good summary if the profiles are parallel. If they are not parallel, the average may not adequately represent the shape of a subset of profiles.

- To deal with this issue, one can consider clustering the profiles and calculating the mean separately for each cluster.

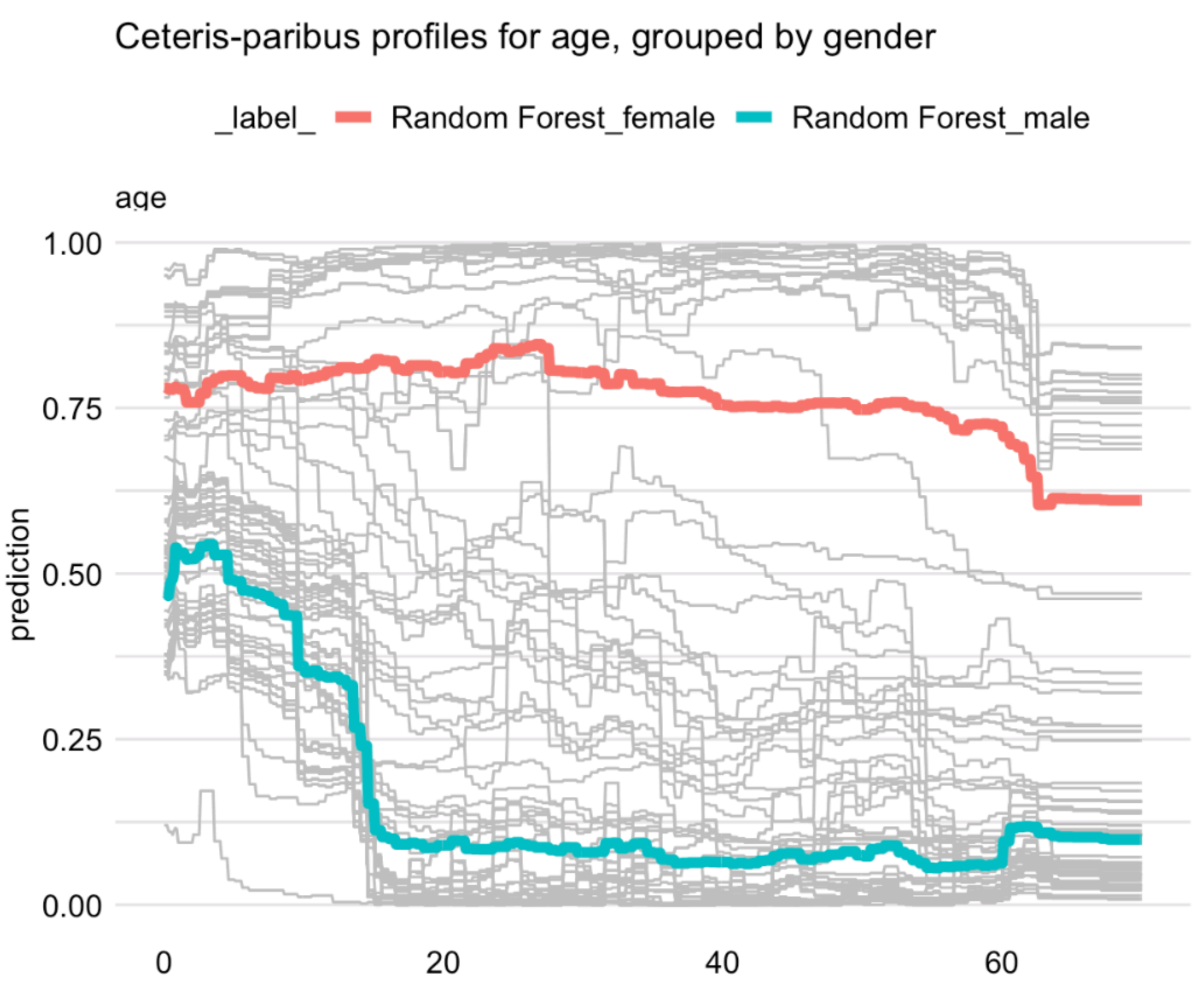
- To cluster the CP profiles, one may use standard methods like K-means or hierarchical clustering. The similarities between observations can be calculated based on the Euclidean distance between CP profiles.

- The plot itself does not allow to identify the variables that may be linked with these clusters, but the additional exploratory analysis could be performed for this purpose.



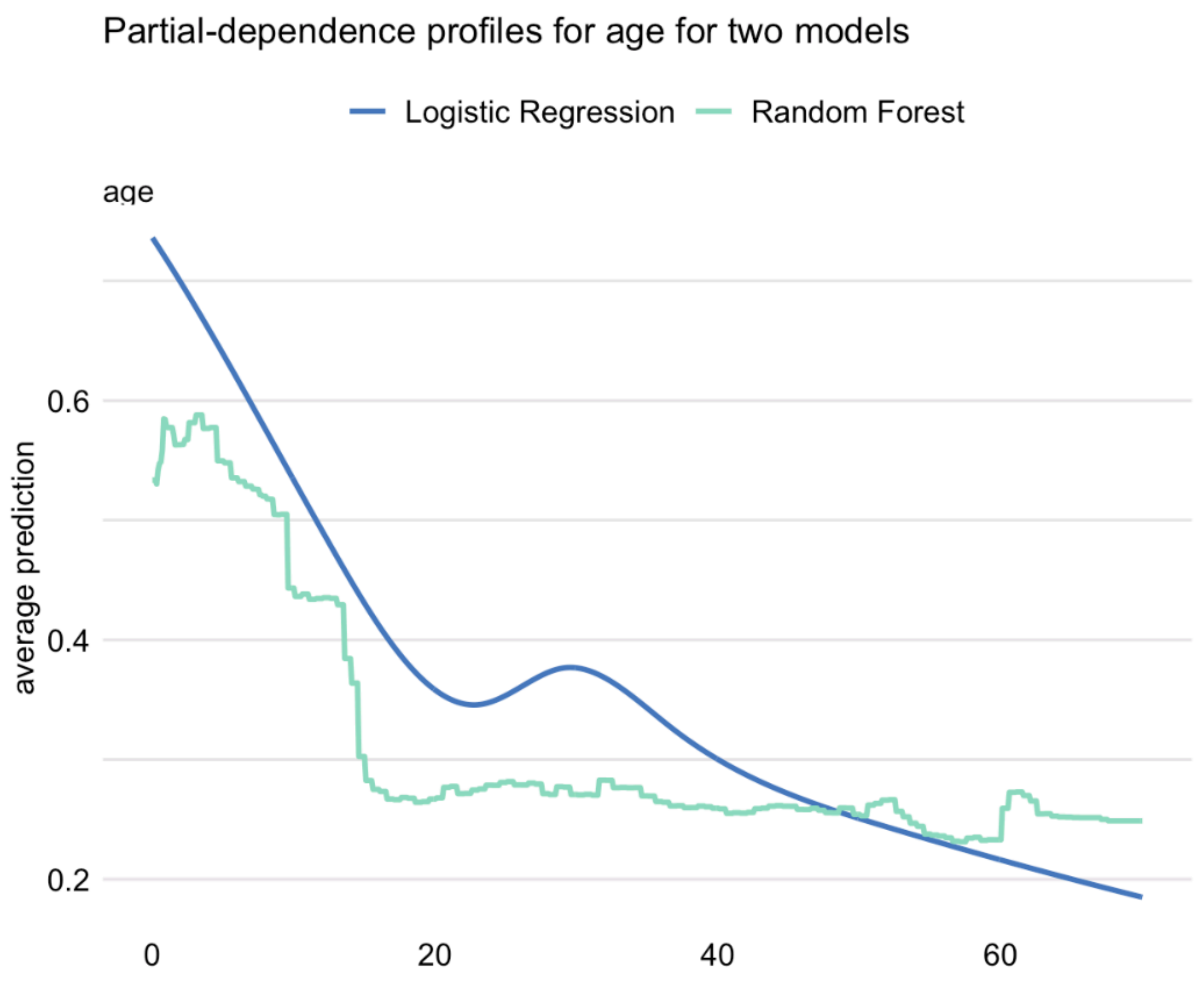
Grouped partial-dependence profiles

- It may happen that we can identify an explanatory variable that influences the shape of CP profiles for the explanatory variable of interest.
- The most obvious situation is when a model includes an interaction between the variable and another one.
- In that case, a natural approach is to investigate the PD profiles for the variable of interest within the groups of observations defined by the variable involved in the interaction.
- The gender-specific averages have different shapes: the predicted survival probability for females is more stable across different ages, as compared to males. Thus, the PD profiles clearly indicate an interaction between age and gender.



Contrastive partial-dependence profiles

- PD profiles can be used to compare between different models.
- The profiles are similar with respect to a general relationship between *age* and the predicted probability of survival (the younger the passenger, the higher chance of survival).
- However, the profile for the random forest model is flatter.
- The difference between both models is the largest at the left edge of the age scale. This pattern can be seen as expected because random forest models, in general, shrink predictions towards the average and they are not very good for extrapolation outside the range of values observed in the training dataset.



Pros and Cons

+ offer a simple way to summarize the effect of a particular explanatory variable on the dependent variable.

+ can be obtained for sub-groups of observations and compared across different models.

- as CP profiles are problematic for correlated explanatory variables, PD profiles are also not suitable for that case.

References

Biecek, Przemyslaw. 2018. "DALEX: Explainers for complex predictive models in R." *Journal of Machine Learning Research* 19 (84): 1–5. http://jmlr.org/papers/v19/18-416.html.

Friedman, Jerome H. 2000. "Greedy Function Approximation: A Gradient Boosting Machine." *Annals of Statistics* 29: 1189–1232.

Greenwell, Brandon M. 2017. "Pdp: An R Package for Constructing Partial Dependence Plots." *The R Journal* 9 (1): 421–36. https://journal.r-project.org/archive/2017/RJ-2017-016/index.html.

Jiangchun, Li. 2018. *Python Partial Dependence Plot Toolbox*. https://pypi.org/project/PDPbox/.

Molnar, Christoph, Bernd Bischl, and Giuseppe Casalicchio. 2018. "iml: An R package for Interpretable Machine Learning." *Journal of Open Source Software* 3 (26): 786. https://doi.org/10.21105/joss.00786.

REMINDER

The lecture at 10:00 includes
a meeting with people from the career office
and there will be some informations about
choosing the best career path for you
and about writing a resume.

Please feel free to send e-mail about your questions!

mustafa.cavus@pw.edu.pl